Three Perspectives on Collaborative Knowledge Acquisition in e-Science

David De Roure and Jeremy Frey
University of Southampton
Schools of ECS and Chemistry
Southampton, SO17 1BJ, UK
dder@ecs.soton.ac.uk, j.g.frey@soton.ac.uk

Abstract

Through a series of e-Science projects we have explored the creation of a complete digital chain of knowledge from the scientific laboratory through to scholarly research output. In this paper we describe this experience and we discuss three perspectives on collaborative knowledge acquisition within the context of this cyberinfrastructure: Publication at Source, Record and Reuse, and Annotation.

1 Introduction

The deluge of data from new experimental techniques and high throughput data acquisition technologies creates important new opportunities for scientific knowledge acquisition. This is a key motivation for e-Science [Hey and Trefethen, 2003] and hence the development of cyberinfrastructure to support the scientific discovery process, with application across a broad range of disciplines.

While individual projects have focused on particular aspects of handling the data deluge or specific parts of the scientific process, when we step back we see a larger picture of scientific knowledge acquisition and its lifecycle. Through a series of e-Science projects in the chemistry domain we have explored the means of creating a complete digital chain of knowledge from the scientific laboratory through to scholarly research output – and back.

In this paper we provide an overview of the key aspects of the systems we have built and we present three perspectives on collaborative knowledge acquisition within the context of this cyberinfrastructure. These perspectives capture the evolving design of our systems over the course of four years based on experience with users, researchers and devlopers.

In Section 2 we set the scene with a holistic view of the scholarly knowledge cycle. We then in Section 3 start in the laboratory, move onto the processing, assimilation and analysis of the data in the e-Science research environment, and finally consider the developments possible in scholarly publishing and dissemination of the data. In Section 4 we look at capturing knowledge in experiments in real-time and in meetings, an essential but often ignored aspect of the scientific discourse when considering the digital arena. Reflecting on this, in Section 5 we present three perspectives on collaborative knowledge acquisition within the context of this cyberinfrastructure: Publication at Source, Record and Reuse, and Annotation. We close with a discussion in Section 6 and conclusions in Section 7.

2 The Scholarly Knowledge Cycle

For many e-Science projects the end result is new data that has been produced faster or which would not have been produced at all without the new techniques. However the overall aim of e-Science is new scientific discovery, and we need to think not just about how to do the experiment better but how we get to the experiment in the first place. This means we need to look at the creation of new scientific knowledge through the overall scientific process – the scholarly knowledge lifecycle, as depicted in figure 1. The data and publication outputs of the scientific process feed into repositories, archives and digital libraries; they are used by researchers and also by learners. Ultimately the cycle is about the flow of knowledge from the laboratory to the scholarly output and back.

The CombeChem project (www.combechem.org) is broadly characteristic of many e-Science projects focusing on using grid techniques to cope with the data deluge, in this case from parallel experiments and high throughput screening or even just many laboratories contributing individually to the global scale chemistry activity. It has focused on gathering data in laboratories and from instruments on the grid, and enabling researchers to use it (for example, by performing compute-intensive computations) to generate results and papers – the scholarly research output. With the related projects eBank and CoAKTinG, CombeChem has set out to explore the broader scholarly lifecycle by looking at interlinking of information from the laboratory through to the scholarly output, which in turn is used by researchers and – through the e-Malaria project [Gledhill et al., 2006] – learners.

Underlying this is the crucial observation that the details of the origins of data are just as important to understanding as their actual values. Hence the CombeChem vision was motivated by the notion of " Publication at Source", a term which describes the need to capture data and its context from the outset and maintain a complete end-to-end connec-

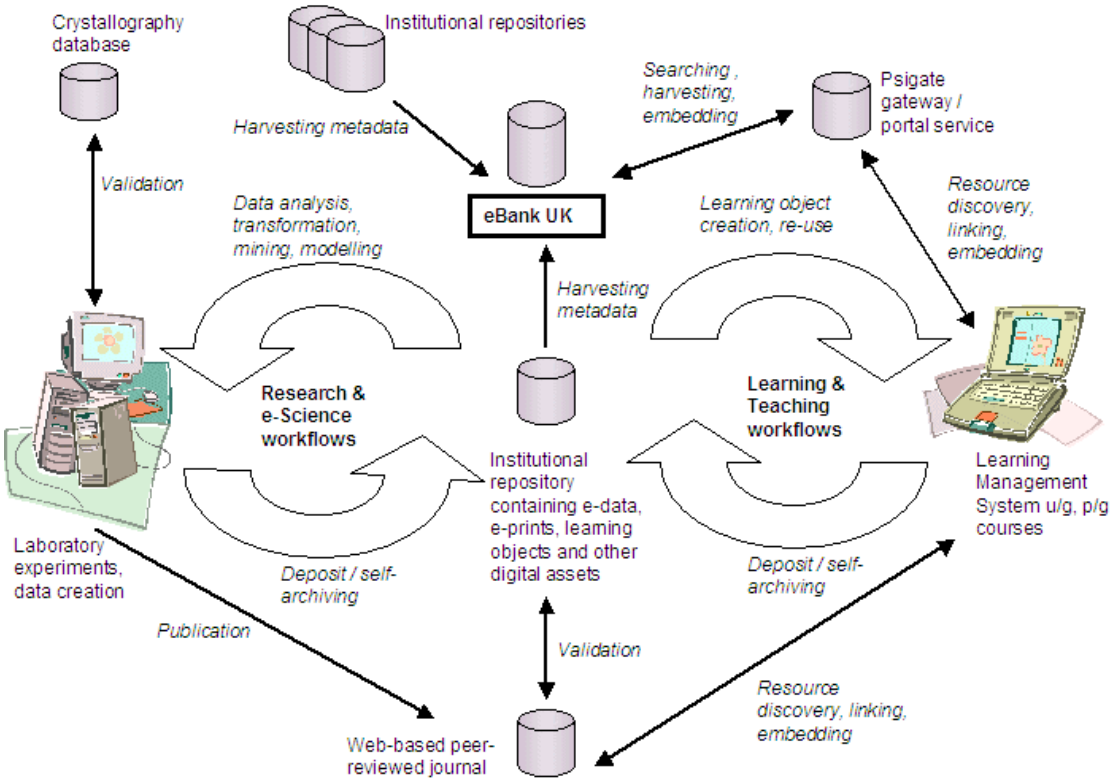

Figure 1: The Scholarly Knowledge Cycle (Liz Lyon, Ariadne Issue 36, July 2003, see www.ariadne.ac.uk)

tion between the laboratory bench and the intellectual chemical knowledge that is published as a result of the investigation [Frey et al., 2002]. It is much easier to collect this information at source and at creation, than attempt to add it later.

The creation of original data is accompanied by information about the experiment and experimental conditions in which it is created. There then follows a chain of processing such as aggregation of experimental data, selection of a particular data subset, statistical analysis, or modelling and simulation. The handling of this information may include annotation of a diagram or editing of a digital image. All of this generates secondary data, accompanied by the information that describes the process that produced it, and this may be maintained in a variety of distinct datastores. Through the principle of publication at source, all this data is made available for subsequent reuse in support of the scientific process, subject to appropriate access control.

Some of these ideas are also demonstrated in the World Wide Molecular Matrix [Murray-Rust, 2002] and the Collaboratory for Multiscale Chemical Science (CMCS) [Myers et al., 2005]. The use of knowledge technologies within the Grid context is explored through many Semantic Grid projects (see semanticgrid.org), notably the myGrid project [Goble et al., 2003] which provides a comprehensive treatment of provenance capture in the context of in silico workflows [Zhao et al., 2004]. CombeChem has become established as a Semantic DataGrid [Taylor et al., 2005].

3 From laboratory to publication

3.1 Design principles

We took a decision at the outset that our research projects would be conducted 'in the wild', i.e. real users actually using our solutions within their scientific work. Hence we took an important decision to base our systems on established practice in the first instance, as this was the most effective route to adoption.

Our design approach adopted five principles [Taylor et al., 2006]:

1. Grounding in established operational practice – our starting point was to study chemists at work;
2. Capturing a rich set of associations between all types of things, expressed pervasively in RDF and hence explicitly addressing the sharing of identifiers;
3. Metadata capture should be automated as far as possible – our goal is efficient augmentation not disruption;
4. Information will be reused in both anticipated and unanticipated ways;
5. The storage, maintenance and transport of metadata will be given equal consideration to data, ensuring availability of accurate metadata, a dependable provenance record and comprehensive understanding of the context of data.

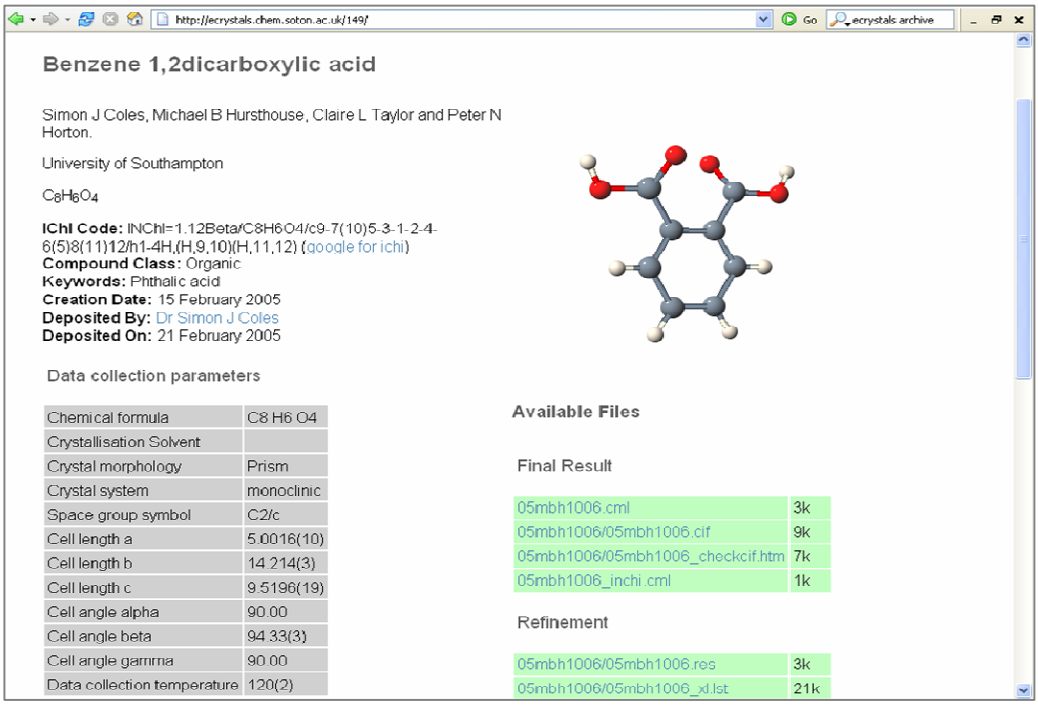

Figure 2: The ecrystals interface

3.2 Knowledge acquisition in the laboratory

The system supports the chemist through the whole life-cycle of an experiment, which we break down into four parts, with the "PPPP" mnemonic: Plan, Perform, Ponder and Publish. Although simplistic, this does capture many of the aspects of the discovery process.

The acquisition starts with planning and performing, using the smart laboratory and Grid-enabled instrumentation [Hughes et al., 2004]. By studying chemists within the laboratory, Electronic Laboratory Notebook technology has been introduced to facilitate the information capture at this earliest stage [schraefel et al., 2004]. Additionally pervasive computing devices are used to capture live metadata as it is created at the laboratory bench, relieving the chemist of the burden of metadata creation [Frey, 2004, Robinson et al., 2005]. This aspect, which is a significant enabler for Publication at Source, is set to grow considerably as pervasive computing deployment advances.

It is significant that this capture makes use of both a record of the researcher's planned activity and what actually occurs. In the UK the chemist has to produce a plan of the experiment as a list of the reagents to be used, and any associated hazards, as part of the COSHH (Control Of Substances Hazardous to Health) assessment.

The plan is a key part of the knowledge capture in support of the publication at source model. It also enables the Electronic Laboratory Notebook to provide a guide to the experiments in the laboratory. Capturing the experiment as a re-usable digital artifact also facilitates sharing and re-use of experiment design.

3.3 Supporting the pondering phase

Experimental results are then used by researchers within CombeChem's grid-based information-and knowledge-sharing environment that integrates existing chemical knowledge, specifically structure and property data sources. The research that is conducted, which may involve simulations using the Grid, leads to new results and to publication.

At the outset we anticipated that we would support this environment by using RDF as a means of integrating across the many established data sources, which include relational databases and third party information providers. This is a good example of the use of RDF triplestores in conjunction with database solutions.

In practice we found that the chemistry researchers were keen to import chemical information directly into the RDF stores. The benefits were the uniform description and the flexible schema afforded by this approach, contrasting the diversity of relational databases where changing schema was impossible or achievable only at very high cost.

This triplestore contains tens of millions of RDF triples and represents a substantial Semantic Web deployment. The chemical data was obtained from a range of publicly available databases including the ZINC database [Irwin and Shoichet, 2005], the National Institutes for Health (NIH) and in particular the National Cancer Institute (NCI) chemical data. We used the open source 3store triplestore software, which was used in a similar harvesting role in the CS AKTive Space project [Shadbolt et al., 2004].

The current target is 200 million triples. However we have moved away from managing everything in one scal-

able triplestore. This harvesting and hoarding approach to the "mash-up" benefits the immediate users but itself is not in the spirit of CombeChem's open approach to publishing knowledge. Rather, we prefer to make the knowledge sources available in RDF and to import them into triplestores as required – the Web itself then becomes the scalable triplestore, and the sources are available for re-use.

The ontologies that were created to support this environment are described in [Taylor et al, 2006].

3.4 Publishing

As an example of the output of this process, the ecrystals interface shown in figure 2 provides a web page complete with a 3D visualisation of a molecule, data collection parameters and links back to the files of data which led to this output (see ecrystals.chem.soton.ac.uk). Behind this simple interface there is a complex picture including a diverse set of stakeholders – the federation model involves data collection, data curation and preservation in databases and databanks, institutional data repositories, aggregator services, portals and publishers. We also produced an academic paper in this form [Rousay et al., 2005].

The interlinking of research data and research publication is the subject of the eBank project, which provides open access crystallography data interlinked with its derived research publications – it is possible to chase back to see exactly where results have come from, or even to find research publications arising from data. In line with the digital library context for this work, OAI (Open Archives Initiative) metadata is harvested from institutional data repositories.

As part of this exercise, the Repository for the Laboratory (R4L) project is developing digital data and document repositories for laboratory-based science (see r4l.eprints.org). R4L addresses the interactions between repositories of primary research data, the laboratory environment in which they operate, and repositories of research publications they feed into.

4 Collaborative Tools

The scholarly knowledge cycle as we have discussed it so far is about sharing and collaboration of an asynchronous nature – publishing things at each other. Through the CoAKTinG project (Collaborative Advanced Knowledge Technologies in the Grid – see www.aktors.org/coakting) we have also addressed synchronous collaboration, as occurs in the meetings that pervade the life of almost all researchers, increasingly taking the form of telephone and videoconferences amongst geographically dispersed colleagues. The CoAKTinG project, which was part of the Advanced Knowledge Technologies Interdisciplinary Research Collaboration, investigated the use of knowledge technologies to enhance meetings, and this has fed through directly into the followon Memetic project (see www.memetic-vre.net). The tools are illustrated in Figure 3.

The objective of CoAKTinG was to advance the state of the art in collaborative mediated spaces for distributed e-Science through the novel application of advanced knowledge technologies [Bachler et al., 2003]. It comprises four tools: meeting capture and replay, instant messaging and presence notification (BuddySpace), graphical meeting and group memory capture (Compendium) and intelligent 'todo' lists (Process Panels). These are integrated through exchanging and storing events according to a set of CoAKTinG ontologies. CoAKTinG conducted case studies with CombeChem, to put the tools in the hands of e-Scientists.

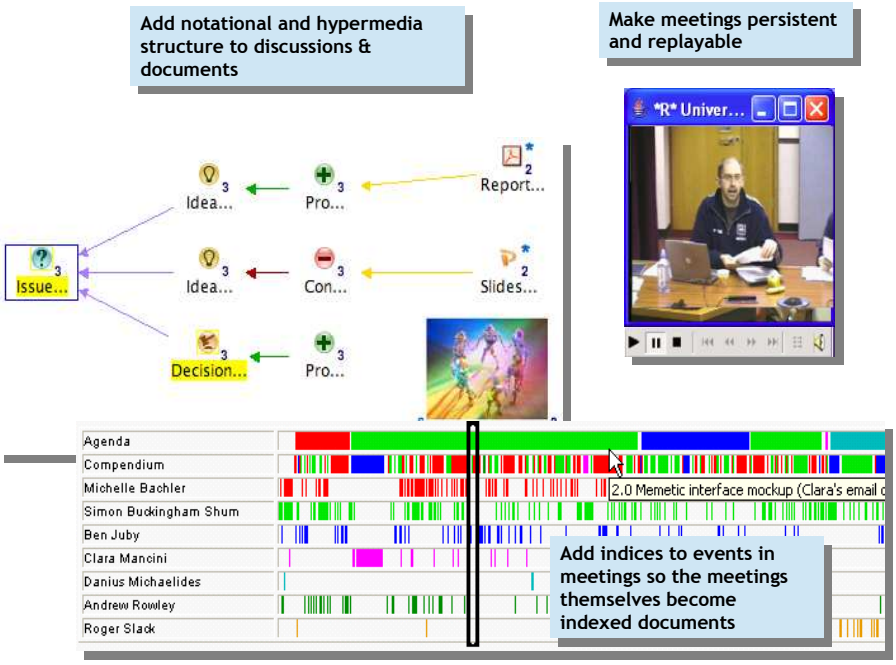

Figure 3: The Memetic Tools (Mike Daw, see www.memetic-vre.net)

The discussions between the two teams led to two notions of integration: "shallow", where the tools are deployed as they are, and "deep" where the tools are more intimately integrated with the CombeChem systems. The BuddySpace system can be adapted to show and track the interactions between staff and equipment, Compendium provides the harness to ensure more adequate capture of the discussions in analysis, while Process Panels provide the means to track the workflow in the performing and pondering phases.

By providing tools for mapping and recording meetings or experiments, CoAKTinG and Memetic make these events persistent and replayable – they turn them into artifacts for retrospective use. In fact they blur the synchronous-asynchronous distinction by effectively turning meetings into documents, which can then be stitched in with the experimental data and results to provide a completely interlinked digital record in line with Publication at Source.

5 Perspectives

As we have moved from requirements – the Publication at Source model – to design, implementation and evaluation, our model for thinking about the systems has evolved and our view on the nature of what needs to be recorded has developed. From concern with the data and its provenance has evolved to deal more directly with the processes that produce the data and their context. This requires the ability to share an even wider range of data and descriptions of process, and be able to manipulate them both, separately and together.

5.1 Publication at Source

This was the original notion that motivated this set of projects – the idea that instead of reading values in a table or graph in an academic paper, scientists need access to the data, derived data resulting from analysis, and the exact context of its creation – tied together to provide effective provenance which is capable of authenticating location, validity and process. In some applications, provenance is required for regulatory purposes, but here our principal goal has been to facilitate re-use of experimental results.

Hence Publication at Source requires two things: rich acquisition and a record of contextual provenance. The latter can be represented as a hypertext. This is demonstrated by the ecrystals interface (Figure 2), in which the provenance record can be navigated in the Web browser to chase back to source data via the stages of analysis. The hypertext can be thought of as a separable structure in its own right (providing navigation through the analysis workflow). As re-use occurs, we can in principle enrich this hypertext.

The process that developed the data here is generally agreed in the community. However in other areas of Chemistry this cannot be taken for granted and the process needs to be explicitly recorded (as a plan and the implementation of the plan) to be exchanged as part of the context.

As Vannevar Bush wrote [Bush, 1945] "It is exactly as though the physical items had been gathered together from widely separated sources and bound together to form a new book. It is more than this, for any item can be joined into numerous trails".

5.2 Record and Reuse

In all the examples given so far we are making recordings – be it data from an X-Ray diffractometer, a video of a performance or entries in the Electronic Lab Notebook – and then reusing this archived digital record in ways that may have been anticipated at the time of capture or, significantly, may not have been, so we ensure that the record is minimally restrictive on re-use. This Record and Reuse perspective underlies e-Science. It is also the basis of our implementation of Publication at Source.

The federation model behind ecrystals illustrates the complexity of the service provision, involving many parties sharing information. Similarly the harvesting of external chemical information into a central store, then the rethink in terms of publishing rather than hoarding, illustrates a shift in perspective towards an open system with a diverse set of knowledge sources which can be assembled for the task at hand. This is consistent with the 'Service Oriented Knowledge Utility' vision of the Next Generation Grid [De Roure, 2006].

The emphasis is on not only the publication of scholarly output which is interlinked to the original data, but on capturing the digital record in all its forms and making this available for re-use. We believe this approach maximises the ability for the information to be reused in ways that were not anticipated at the outset.

5.3 Annotation

In a sense, Annotation is our implementation of Record and Reuse. The capture of contextual information at source is annotation upon the data values; the provenance record is an annotation of the data. In other words, to implement Record and Reuse we are building an annotation infrastructure – one that handles multiple distributed interlinked annotations, and which again supports re-use.

In this perspective we see the need for creation and maintenance of annotations – it makes explicit the fact that we have a metadata lifecycle and infrastructure which is deeply associated with the data infrastructure but can be viewed as a distinct infrastructure with its own set of engineering demands.

We believe that annotation is key to making data reusable, by creating a suitably annotated form of digital record, but also that it is key to the process of re-use – by creating further annotation when data is re-used. All use of our published knowledge is effectively annotation upon it and may add value. While much of our annotation is produced automatically, there is also a role for explicit annotation by users – for example we have explored the blogging of experiment data.

Vannevar Bush also wrote "There is a growing mountain of research. But there is increased evidence that we are being bogged down today as specialization extends. The investigator is staggered by the findings and conclusions of thousands of other workers—conclusions which he cannot find

time to grasp, much less to remember, as they appear. Yet specialization becomes increasingly necessary for progress, and the effort to bridge between disciplines is correspondingly superficial."

To benefit from these effects we need to share the knowledge that users are creating. Projects like R4L provide tooling for this within the research environment, addressing the interactions between repositories of primary research data, the laboratory environment in which they operate and repositories of research publications into which they ultimately feed (through documented interpretation and analysis of the results and in explicit linking and citation of the data sets). If we can create a common shared space then we can achieve the benefits of the scale of usage.

The myGrid (see myGrid.org.uk) e-Science project has embraced the annotation perspective [Goble et al., 2006] and is focusing on collaboration through the creation of myExperiment, a collaborative platform for life scientists to share experimental information – in this case *in silico* workflows.

Additional to the direct benefits to the users, collecting and sharing information about what people are actually doing with the cyberinfrastructure provides a basis for research into enhancing the research environment. This includes the performance and function of the system, but also the usability – again, users are key.

6 Discussion - The Grid of people

The three perspectives explain our model of collaborative knowledge acquisition in e-Science. Publication at Source describes our model of publishing reusable knowledge, in the form of scholarly publication output and accompanying contextualised data. Record and Reuse is how we achieve this, realised within an open distributed environment with multiple federated stakeholders. The annotation perspective describes how the knowledge is acquired and maintained through users participating in this environment.

Although the knowledge and data are deeply entwined, we can also look at the degree to which the metadata can be viewed as a separable infrastructure – it certainly has its own lifecycle, owners etc. This is particularly evident when we think of the provenance hypertext in Publication at Source, and the accumulation of annotation in the Annotation perspective. Our real experience of using these systems is teaching us about this second infrastructure.

Dan Atkins, the Director of the NSF Office of Cyberinfrastructure, uses a picture of three symmetric interlocking rings ('Borromean rings') to illustrate the alignment of endeavours necessary to create, provision, and apply cyberinfrastructure to enhance the activities of knowledge communities. Removing any one of the three symmetric rings destroys the synergy:

- Transformative Application – to enhance discovery & learning;
- Provisioning – Creation, deployment and operation of advanced cyberinfrastructure;
- R&D to enhance technical and social dimensions of future cyberinfrastructure systems.

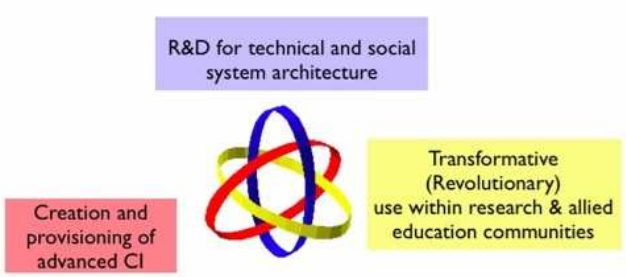

Figure 4. Borromean Rings, after Atkins

The third point is very much the subject of our discussion here. It is easy to think about cyberinfrastructure with a service provision mentality, and to focus on connecting people to resources. We suggest instead that the many users of our knowledge sources are all participants in the Grid – not just consumers of information but generators of new content of value to others. With the appropriate tools we are beginning to see a deeper sense of people contributing to the cyberinfrastructure through being participants rather than just consumers. It is as if we are taking the social mechanism of the scholarly knowledge cycle – publishing and reusing – and applying it to the collaborative knowledge acquisition in the cyberinfrastructure.

7 Conclusion

The key points arising from our reflections on six years of developing and using systems for collaborative knowledge acquisition in e-Science are as follows:

1. *Holistic View.* Accelerating time-to-discovery means accelerating time-to-experiment, not just accelerating the experiment and subsequent data-processing;
2. *Publishing data, results and experiments*. As much attention should be paid to publishing knowledge as acquiring it. This is how we create an open environment of federated information sources to support the ongoing process of scientific discovery and development;
3. *Making data, results and experiments re-usable*. Capturing context of data and provenance of results is essential for flexible re-use of results. We have shown how Semantic Web technologies can be used to maintain the provenance record;
4. *Metadata Lifecycle.* As illustrated by the Annotation perspective, we are effectively creating a second information infrastructure with its own lifecycle, closely entwined with our experiments and data;
5. *Collaboration*. People are participants, not just consumers, and knowledge of all forms – including experiments themselves – can be shared and re-used.

It is interesting to ask whether these principles are unique to the chemistry domain or can be applied elsewhere. We suggest that they are generic and look forward to further exploration to investigate this.

Acknowledgments

The authors wish to thank the CombeChem, eBank and CoAKTing projects and teams, as well as R4L and Memetic. These projects were funded by the UK e-Science programme and by JISC. We also thank our colleagues in the e-Science community, especially the [my]Grid project, and the Advanced Knowledge Technologies and Equator IRCs.

References

[Bachler et al., 2003] M. Bachler, S. Buckingham-Shum, J. Chen-Burger, J. Dalton, D. De Roure, M. Eisenstadt, J. Frey, J. Komzak, D. Michaelides, K. Page, S. Potter, N. Shadbolt and A. Tate, *Chain ReAKTing: Collaborative Advanced Knowledge Technologies in the CombeChem Grid*, in S. Cox, ed., *UK e-Science All Hands Meeting*, EPSRC, Nottingham, UK, 2003.

[Bush, 1945] V. Bush, *As We May Think*, The Atlantic Monthly, 176 (1945), pp. 101-108.

[De Roure, 2006] D. De Roure, *Future for European Grids: GRIDs and Service Oriented Knowledge Utilities*, EU Grid Technologies Unit, 2006.

[Frey, 2004] J. G. Frey, *Dark Lab or Smart Lab: The challenges for 21st century laboratory software*, Organic Process Research & Development, 8 (2004), pp. 1024-1035.

[Frey et al., 2002] J. G. Frey, D. De Roure and L. A. Carr, *Publication At Source: Scientific Communication from a Publication Web to a Data Grid, Euroweb 2002 Conference, The Web and the GRID: from e-science to e-business*, BCS, Oxford, UK, 2002.

[Gledhill et al., 2006] R. Gledhill, S. Kent, B. Hudson, W. G. Richards, J. W. Essex and J. G. Frey, *A Computer Aided Drug Discovery System for Chemistry Teaching*, Journal of Chemical Information and Modeling, 46 (2006), pp. 960-970.

[Goble et al., 2006] C. Goble, O. Corcho, P. Alper and D. D. Roure, *e-Science and the Semantic Web: A Symbiotic Relationship*, in N. Lavrac, L. Todorovski and K. P. Jantke, eds., *Discovery Science, 9th International Conference, DS 2006*, Springer, Barcelona, Spain, 2006, pp. 1-12.

[Goble et al., 2003] C. A. Goble, S. Pettifer and R. Stevens, *Knowledge Integration: In silico Experiments in Bioinformatics*, in I. Foster and C. Kesselman, eds., *The Grid: Blueprint for a New Computing Infrastructure Second Edition*, Morgan Kaufman, 2003.

[Hey and Trefethen, 2003] A. J. G. Hey and A. E. Trefethen, *The Data Deluge: An e-Science Perspective*, in F. Berman, G. C. Fox and A. J. G. Hey, eds., *Grid Computing - Making the Global Infrastructure a Reality*, Wiley and Sons, 2003, pp. 809-824.

[Hughes et al., 2004] G. Hughes, H. Mills, D. D. Roure, J. G. Frey, L. Moreau, m. c. schraefel, G. Smith and E. Zaluska, *The Semantic Smart Laboratory: A system for supporting the chemical e-Scientist*, Organic & Biomolecular Chemistry., 2 (2004), pp. 3284-3293.

[Irwin and Shoichet, 2005] J. J. Irwin and B. K. Shoichet, *ZINC - a free database of commercially available compounds for virtual screening*, J. Chem. Inf. Model., 45 (2005), pp. 177-82.

[Murray-Rust, 2002] P. Murray-Rust, *The World Wide Molecular Matrix - a peer-to-peer XML repository for molecules and properties, Euroweb 2002 Conference, The Web and the GRID: from e-science to e-business*, BCS, Oxford, UK, 2002, pp. 163-164.

[Myers et al., 2005] J. D. Myers, T. C. Allison, S. Bittner, B. Didier, M. Frenklach, J. William H. Green, Y.-L. Ho, J. Hewson, W. Koegler, C. Lansing, D. Leahy, M. Lee, R. McCoy, M. Minkoff, S. Nijsure, G. v. Laszewski, D. Montoya, C. Pancerella, R. Pinzon, W. Pitz, L. A. Rahn, B. Ruscic, K. Schuchardt, E. Stephan, A. Wagner, T. Windus and C. Yang, *A Collaborative Informatics Infrastructure for Multi-Scale Science*, Cluster Computing Special Issue: Challenges of Large Applications in Distributed Environments, 8 (2005).

[Robinson et al., 2005] J. M. Robinson, J. G. Frey, A. D. Reynolds, B. V. Bedi and A. J. Stanford-Clark, *From the laboratory to the mobile phone: Middleware for laboratory data acquisition using the publication subscribe model, e-Science 2005, Workshop on Instruments and Sensors on the Grid*, IEEE Computer Society, Melbourne, Australia, 2005, pp. 561-569.

[Rousay et al., 2005] E. R. Rousay, H. Fu, J. M. Robinson, J. W. Essex and J. G. Frey, *Grid-based dynamic electronic publication: a case study using combined experiment and simulation studies of crown ethers at the air/water interface*, Philosophical Transactions of the Royal Society A: Mathematical Physical and Engineering Sciences, 363 (2005), pp. 2075-2095.

[schraefel et al., 2004] m. c. schraefel, G. V. Hughes, H. R. Mills, G. Smith, T. R. Payne and J. Frey, *Breaking the Book: Translating the Chemistry Lab Book to a Pervasive Computing Environment, Conference on Human Factors in Computing Systems (CHI)*, ACM Press, Vienna, Austria 2004, pp. 25 - 32.

[Shadbolt et al., 2004] N. Shadbolt, N. Gibbins, H. Glaser, S. Harris and m. c. schraefel, *CS AKTive Space, or How We Learned to Stop Worrying and Love the Semantic Web* IEEE Intelligent Systems 19 (2004), pp. 41-47.

[Taylor et al., 2006] K. Taylor, J. W. Essex, J. G. Frey, H. R. Mills, G. Hughes and E. J. Zaluska, *The Semantic Grid and Chemistry: Experiences with CombeChem*, Journal of Web Semantics, 4 (2006), pp. 84-101.

[Taylor et al., 2005] K. Taylor, R. Gledhill, J. W. Essex, J. G. Frey, S. W. Harris and D. D. Roure, *A Semantic Datagrid for Combinatorial Chemistry, 6th IEEE/ACM International Workshop on Grid Computing*, Seattle, 2005, pp. 148-155.

[Zhao et al., 2004] J. Zhao, C. Wroe, C. Goble, R. Stevens, D. Quan and M. Greenwood, *Using Semantic Web Technologies for Representing e-Science Provenance, 3rd International Semantic Web Conference ISWC2004*, Springer, Hiroshima, Japan, 2004.